International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 93 - 97
https://irjaeh.com

Real-Time Sign Language Recognition and Translation Using Deep Learning Techniques

Tazyeen Fathima[1], Ashif Alam[2], Ashish Gangwar[3], Dev Kumar Khetan[4], Prof. Ramya K[5]*

[1,2,3,4]UG, Artificial Intelligence and Machine Learning Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.

[5]Assistant Professor, Artificial Intelligence and Machine Learning Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.

Emails: *20beam055@dsce.edu.in[1], 20beam049@dsce.edu.in[2], 20beam050@dsce.edu.in[3], 20beam051@dsce.edu.in[4], kramya424@gmail.com[5]*

****Corresponding Author Orcid ID:*** *https://orcid.org/0009-0001-2597-1044*

Abstract

Sign Language Recognition (SLR) recognizes hand gestures and produces the corresponding text or speech. Despite advances in deep learning, the SLR still faces challenges in terms of accuracy and visual quality. Sign Language Translation (SLT) aims to translate sign language images or videos into spoken language, which is hampered by limited language comprehension datasets. This paper presents an innovative approach for sign language recognition and conversion to text using a custom dataset containing 15 different classes, each class containing 70-75 different images. The proposed solution uses the YOLOv5 architecture, a state-of-the-art Convolutional Neural Network (CNN) to achieve robust and accurate sign language recognition. With careful training and optimization, the model achieves impressive mAP values (average accuracy) of 92% to 99% for each of the 15 classes. An extensive dataset combined with the YOLOv5 model provides effective real-time sign language interpretation, showing the potential to improve accessibility and communication for the hearing impaired. This application lays the groundwork for further advances in sign language recognition systems with implications for inclusive technology applications.

Keywords: *Sign Language Recognition (SLR), Sign Language Translation (SLT), YOLO V5 architecture, Convolution Neural Network (CNN), mAP values*

1. Introduction

The main form of communication for the deaf and dumb is sign language (SL), which differs from spoken or written language in terms of vocabulary, meaning, and grammar. There are between 138 and 300 distinct forms of sign language used worldwide; in India, where there are 7 million deaf people, there are only around 250 licenced interpreters. It is difficult to teach sign language to the community because of this lack. To overcome communication hurdles, sign language recognition uses computer vision and deep learning to identify hand motions and transform them into text or voice [1]. The importance of a Sign Language Recognition (SLR) system for hard-of-hearing and speech-impaired people is emphasised in the study. Current SLRs frequently depend on several depth sensor cameras or pricey wearable sensors. The suggested method presents a framework for multilingual sign language recognition that is based on vision and tracks and extracts multi-semantic manual co-articulations, such as one- and two-handed signals, in addition to non-manual components like body language and facial emotions. The objective is to isolate and extract different signals and non-manual motions to create a

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 93 - 97
https://irjaeh.com

realistic, multi-signer Indo-Russian Sign Language Database [2]. To close the communication gap between the public and the hearing-impaired, the authors present a Hybrid Deep Neural Architecture (H-DNA) that integrates CNN, LSTM, GRU, and GAN for real-time sign language detection and translation. The H-DNA model highlights how it may improve communication in a variety of sign languages by showcasing accurate detection and translation [3]. The paper explores the difficulties and methods involved in translating text into sign language or in obtaining voice from movies in sign language. Translation, interpretation, and sign segmentation are some of the tasks involved in the process. Identifying sign glosses is a common emphasis of current research instead of doing a whole translation. To improve translation accuracy, the study presents Sign Language Transformers, which use an encoder-decoder architecture with gloss representations. It also gives baseline data and offers possible directions for further study [4]. The paper offers a framework for translating sign language films into spoken language using sign language translation (SLT). Modules for quantifying semantic similarity, conditional sentence construction, and word existence verification are all included in the system. By resolving issues like word order variances and missing words in sign language movies, this system increases the efficacy of SLT. Results from experiments using sign language datasets show how useful the method is for improving communication between the hearing and the deaf groups [5].

2. Experimental Methods or Methodology

2.1 Data Preprocessing

2.1.1 Custom Dataset Creation

To begin the project, we carefully chose a unique dataset designed specifically for sign language interpretation and identification. This carefully selected dataset includes a wide range of hand gestures and emotions that are often used in sign language communication. Interestingly, our dataset includes 70–75 photos for each of 15 different classes. The wide range of sign language expressions that are represented by this selection guarantees a solid basis for our model's training, which will enable it to correctly interpret and convert these gestures into meaningful text or speech [6].

2.1.2 Data Preprocessing and Labelling

We preprocess and label the gathered data by utilizing Roboflow, a potent tool for dataset management. To guarantee consistency and accuracy in training, this stage entails cleaning, normalizing, and annotating the dataset.

2.2 Model Construction

2.2.1 YOLO v5 Model Architecture Selection

We chose the YOLO v5 architecture because it has a well-established track record of being effective in object identification tasks, which makes it a perfect fit for our goals about sign language recognition. "You Only Look Once," or YOLO is well known for its real-time image processing capabilities and ability to forecast bounding boxes around things. The architecture's speed and efficacy are enhanced by its single forward run through the neural network, which precisely matches our objectives for sign language identification [7].

2.2.2 Training the YOLO v5 Model

We next use our carefully chosen and labelled dataset to train the model with the chosen YOLO v5 architecture. Through a series of iterative procedures, the model is trained to identify and classify a wide range of sign language gestures that are included in the dataset. Exposure to varied hand gestures allows the model to fine-tune its parameters, ultimately achieving skill in reliably detecting and categorizing signals in a variety of contexts. System Overview is shown in Figure 1.

2.2.3 Fine-Tuning and Optimization

We go into a crucial stage of optimization and fine-tuning after the first training phase. To improve accuracy and robustness, we tweak hyperparameters as we examine the finer points of the model's performance. Any particular difficulties or problems found during the first training are methodically resolved. By carefully optimizing the model, we want to improve its performance and make it more resilient to various sign language expressions and environmental factors.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 93 - 97
https://irjaeh.com

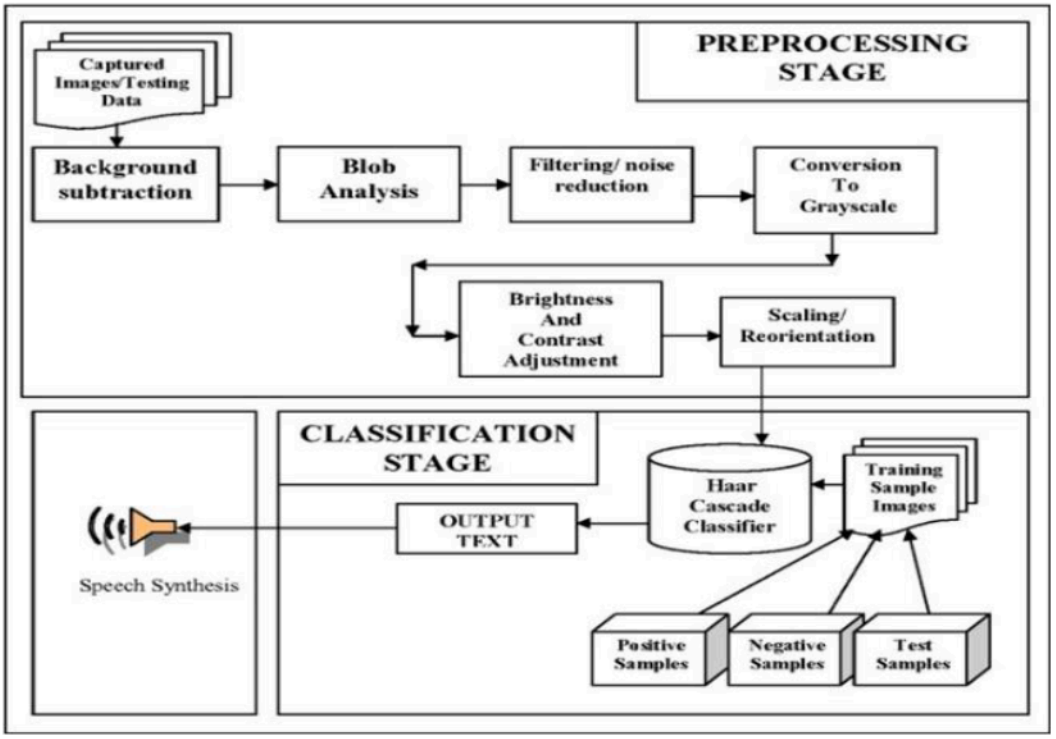

Figure 1 System Overview

2.3 Evaluation and Deployment

2.3.1 Assessment and Validation

To determine the model's capacity for generalization and to guarantee dependable performance in a range of scenarios, a thorough assessment and validation are carried out utilizing distinct datasets.

2.3.2 Deployment Readiness

After the model performs well enough, we get ready to deploy it. This entails drafting a plan for smooth integration and taking into account any deployment-specific issues [8].

2.3.3 User Interface Implementation

The usability of the programme depends on the design of a user-friendly interface. Our goal is to develop a user-friendly interface that supports several forms of communication, such as text, speech, and sign language.

2.3.4 Integration of Accessibility Features

We use features like text-to-speech, screen readers, and hapticfeedback to improve accessibility. The programme is more accessible and easier to use for people with a variety of impairments thanks to these enhancements.

2.3.5 Testing with User Feedback

A wide range of users, including those who are deaf, dumb, or have other impairments, test the completed programme. We actively gather user input to pinpoint areas that need development and enhance the usability of the programme.

2.3.6 Maintenance and Deployment

The programme is placed on a cloud-based platform, guaranteeing its ongoing efficacy and availability. Frequent maintenance is carried out to fix any new problems and provide upgrades or improvements as required.

3. Results and Discussion

The proposed model seamlessly recognizes and converts hand movements into text as shown in Figure 2 it closes social divides and promotes efficient communication. Its precise comprehension of many sign language expressions makes a substantial contribution to inclusive communication, overcoming linguistic barriers, and improving accessibility. Object detection output from the YOLO model shows different signs shown in Figure 2.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 02 February 2024
Page No: 93 - 97
https://irjaeh.com

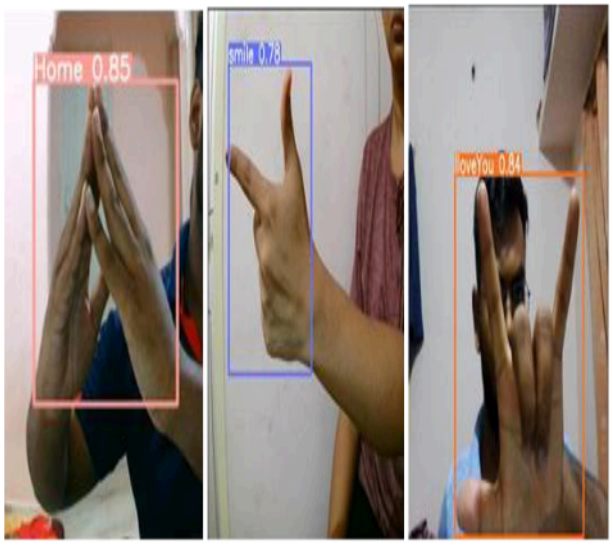

Figure 2 Object detection output from the YOLO model showing different signs

After being trained on 15 different classes, each with 70–75 photos, the model's Mean Average Precision (mAP) metrics shown below in Figure 3 demonstrate its outstanding accuracy. The strong performance, which was achieved by training on a varied dataset, highlights the model's efficacy in promoting inclusive and accessible communication platforms [9].

Class	Images	Instances	P	R	mAP50	mAP50-95:
all	87	87	0.951	0.968	0.959	0.76
Hello	87	4	0.977	1	0.995	0.778
Home	87	5	1	0.92	0.995	0.821
IloveYou	87	7	0.993	1	0.995	0.778
No	87	6	0.981	1	0.995	0.73
Please	87	8	0.983	1	0.995	0.731
Thanks	87	13	0.989	1	0.995	0.82
Yes	87	3	0.972	1	0.995	0.664
all is well	87	5	0.968	1	0.995	0.835
break	87	5	0.973	1	0.995	0.712
help	87	5	0.969	1	0.995	0.9
here	87	6	0.974	1	0.995	0.778
how are you	87	5	0.966	1	0.995	0.846
peace	87	5	0.967	1	0.995	0.863
smile	87	5	0.971	1	0.995	0.826
time	87	5	0.584	0.6	0.453	0.321

Figure 3 mAP values of each class

Furthermore, the confusion matrix's thorough study as in Figure 4 sheds further light on the model's performance. The aforementioned matrix provides valuable insights into the intricate details of identification and translation. It illustrates the model's ability to discriminate among the 15 classes, which contributes to our comprehension of its strengths and potential areas for improvement.

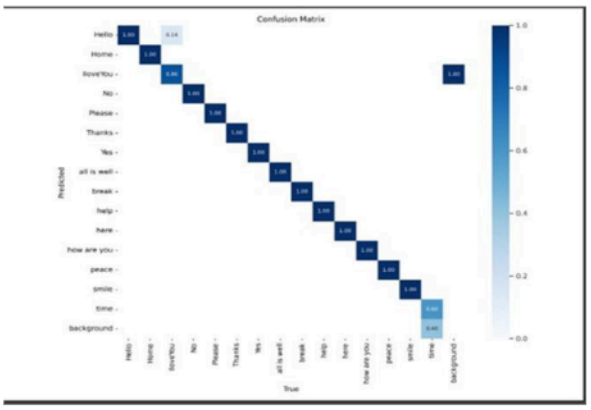

Figure 4 Confusion matrix

Conclusion

The recognition and translation of sign language have entered a new era of advancements because of the incorporation of cutting-edge deep learning techniques, particularly neural networks and transformers. Accurate sign language identification is a crucial difficulty that this revolutionary wave attempts to address. Scholars have worked hard to address the challenges of identifying semantic co-articulations, nonmanual features, and spatial-temporal features present in sign language expressions, demonstrating a commitment to improving the area. Simultaneously, research into cutting-edge translation methods like cross-modal reranking and word existence verification has produced encouraging results, especially when conducted on datasets that are available to the public. These creative methods represent a major advancement in the development of translation techniques by improving the interpretability of sign language statements. Moreover, the terrain has been enhanced by the use of emotion analysis algorithms, improving the accuracy of text and voice classification in sign language scenarios. The DED algorithm and the suggested empathetic speech synthesis approach are examples of pioneering technologies that highlight the practical significance of these breakthroughs. They stress the continued dedication to study, to advance sign language processing, and to promote effective communication in linguistically heterogeneous communities, hence advancing inclusion and accessibility.

References

[1]. Satwik Ram Kodandaram, N Pavan Kumar and Sunil G L, “Sign Language Recognition”, vol.12, No.14 (2021), 994 – 1009. Doi: 10.13140/RG.2.2.29061.47845

[2]. E. Rajalakshmi et al., "Multi-Semantic

Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226-2238, 2023, doi: 10.1109/ACCESS.2022.3233671.

[3]. B. Natarajan et al., "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," in IEEE Access, vol. 10, pp. 104358-104374, 2022, doi: 10.1109/ACCESS.2022.3210543.

[4]. N. Cihan Camgöz, O. Koller, S. Hadfield, and R. Bowden, "Sign Language Transformers: Joint Endto-End Sign Language Recognition and Translation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10020-10030, doi: 10.1109/CVPR42600.2020.01004.

[5]. J. Zhao, W. Qi, W. Zhou, N. Duan, M. Zhou, and H. Li, "Conditional Sentence Generation and Cross-Modal Reranking for Sign Language Translation," in IEEE Transactions on Multimedia, vol. 24, pp. 2662-2672, 2022, doi: 10.1109/TMM.2021.3087006.

[6]. Moganapriya, C., et al. "Dry machining performance studies on TiAlSiN coated inserts in turning of AISI 420 martensitic stainless steel and multi-criteria decision-making using Taguchi-DEAR approach." Silicon (2021): 1-14.

[7]. Kaliyannan, Gobinath Velu, et al. "Development of sol-gel derived gahnite anti-reflection coating for augmenting the power conversion efficiency of polycrystalline silicon solar cells." Materials Science-Poland 37.3 (2019): 465-472.

[8]. Velu Kaliyannan, Gobinath, et al. "An extended approach on power conversion efficiency enhancement through deposition of ZnS-Al 2 S 3 blends on silicon solar cells." Journal of Electronic Materials 49 (2020): 5937-5946.

[9]. Sathishkumar, T. P., et al. "Investigation of chemically treated randomly oriented sansevieria ehrenbergii fiber reinforced isophthallic polyester composites." Journal of Composite Materials 48.24 (2014): 2961-2975.